



US006341373B1

(12) **United States Patent**
Shaw

(10) Patent No.: **US 6,341,373 B1**
(45) Date of Patent: ***Jan. 22, 2002**

(54) **SECURE DATA DOWNLOADING,
RECOVERY AND UPGRADING**

(75) Inventor: **Robert Shaw, Los Altos, CA (US)**

(73) Assignee: **Liberate Technologies, San Carlos, CA (US)**

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **09/728,700**

(22) Filed: **Dec. 1, 2000**

Related U.S. Application Data

(63) Continuation of application No. 09/080,577, filed on May 18, 1998, which is a continuation-in-part of application No. 08/770,238, filed on Dec. 20, 1996, now Pat. No. 5,991,799
(60) Provisional application No. 60/046,749, filed on May 16, 1997.

(51) Int. Cl.⁷ **G06F 9/445; G06F 17/30**

(52) U.S. Cl. **717/11; 709/203; 709/218; 709/222; 713/2**

(58) Field of Search **717/11; 709/221, 709/222, 218, 203; 707/10, 203; 713/2**

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,035,835 A 7/1977 Poetsch
5,436,673 A 7/1995 Bachmann et al.
5,680,458 A 10/1997 Spelman et al.
5,752,042 A * 5/1998 Cole et al. 717/11
5,790,796 A * 8/1998 Sadowsky 709/221

5,808,628 A 9/1998 Hinson et al.
5,867,166 A 2/1999 Myhrvold et al.
5,870,759 A * 2/1999 Bauer et al. 707/201
5,974,461 A * 10/1999 Goldman et al. 709/224
5,991,542 A * 11/1999 Han et al. 717/11
6,006,034 A * 12/1999 Heath et al. 717/11
6,009,274 A * 12/1999 Fletcher et al. 717/11
6,049,628 A 4/2000 Chen et al.
6,049,671 A * 4/2000 Slivka et al. 717/11
6,199,204 B1 * 3/2001 Donohue 717/11
6,202,207 B1 * 3/2001 Donohue 717/11

OTHER PUBLICATIONS

Segal & Frieder, "On-The-Fly Program Modification: Systems for Dynamic Updating," IEEE Software, Mar. 1993, pp. 53-65.*

Hauptmann & Wasel, "On-line Maintenance with On-the-fly Software Replacement," Proceedings of IEEE Third International Conference on Configurable Distributed Systems, May 6-8, 1996, pp. 70-80.*

* cited by examiner

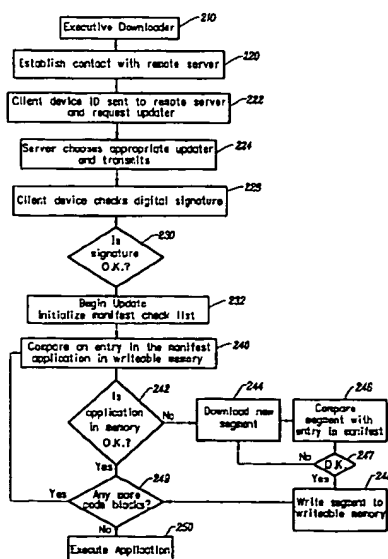
Primary Examiner—Tuan Q. Dam

(74) Attorney, Agent, or Firm—Michael A. Glenn

(57) ABSTRACT

The invention provides an improved method and system for secure downloading, recovery, and upgrading of data. A client device receives information from a server device using a reliable software modules stored in permanent memory in the client device. The reliable software modules perform software and data integrity tests, and locate and retrieve data for recovery or upgrade of the client device. The client device confirms the trustworthiness of the received information device by comparing digital signatures or digests for the information it receives with known digital certificates in the reliable software module.

30 Claims, 3 Drawing Sheets



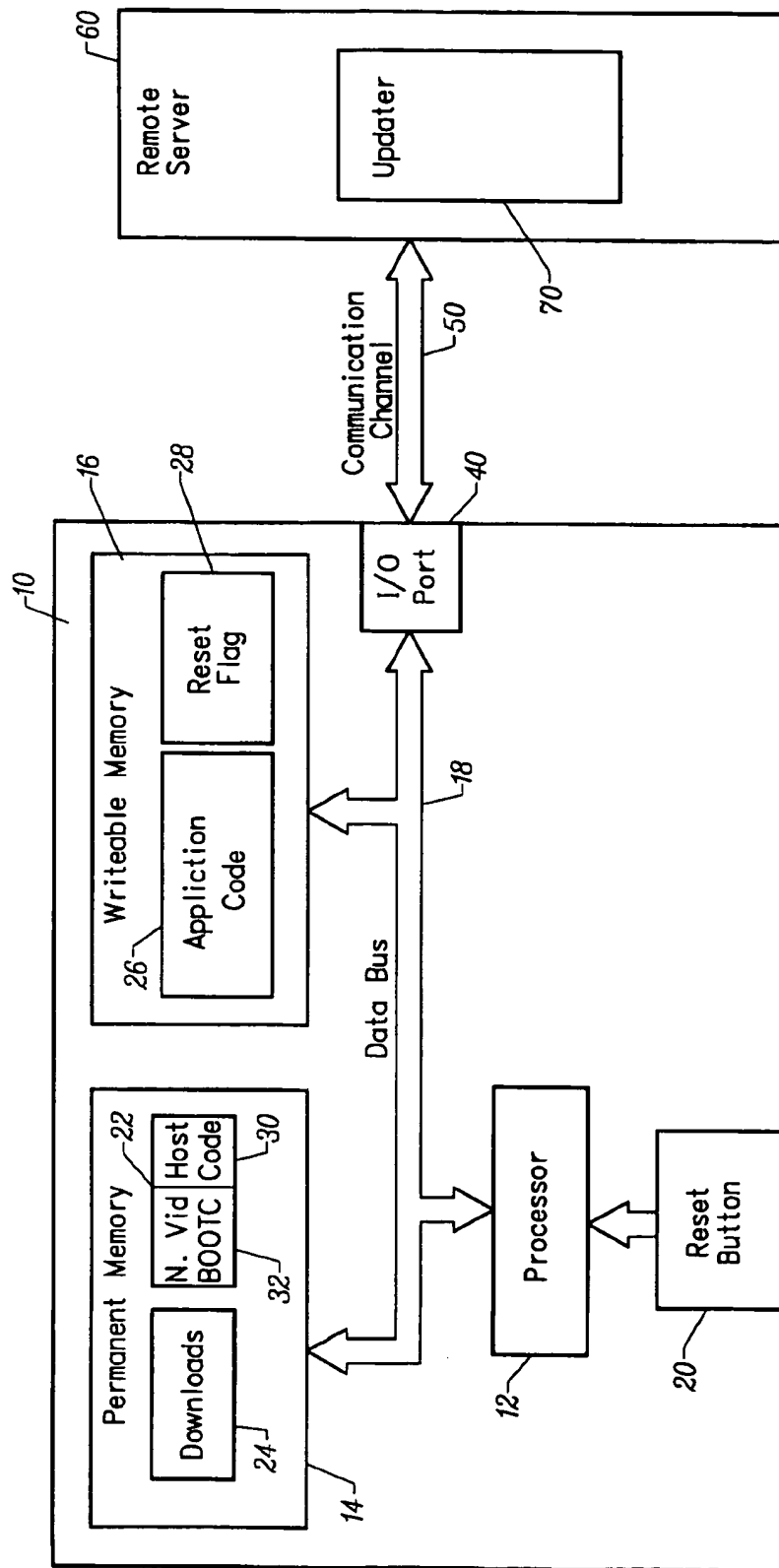
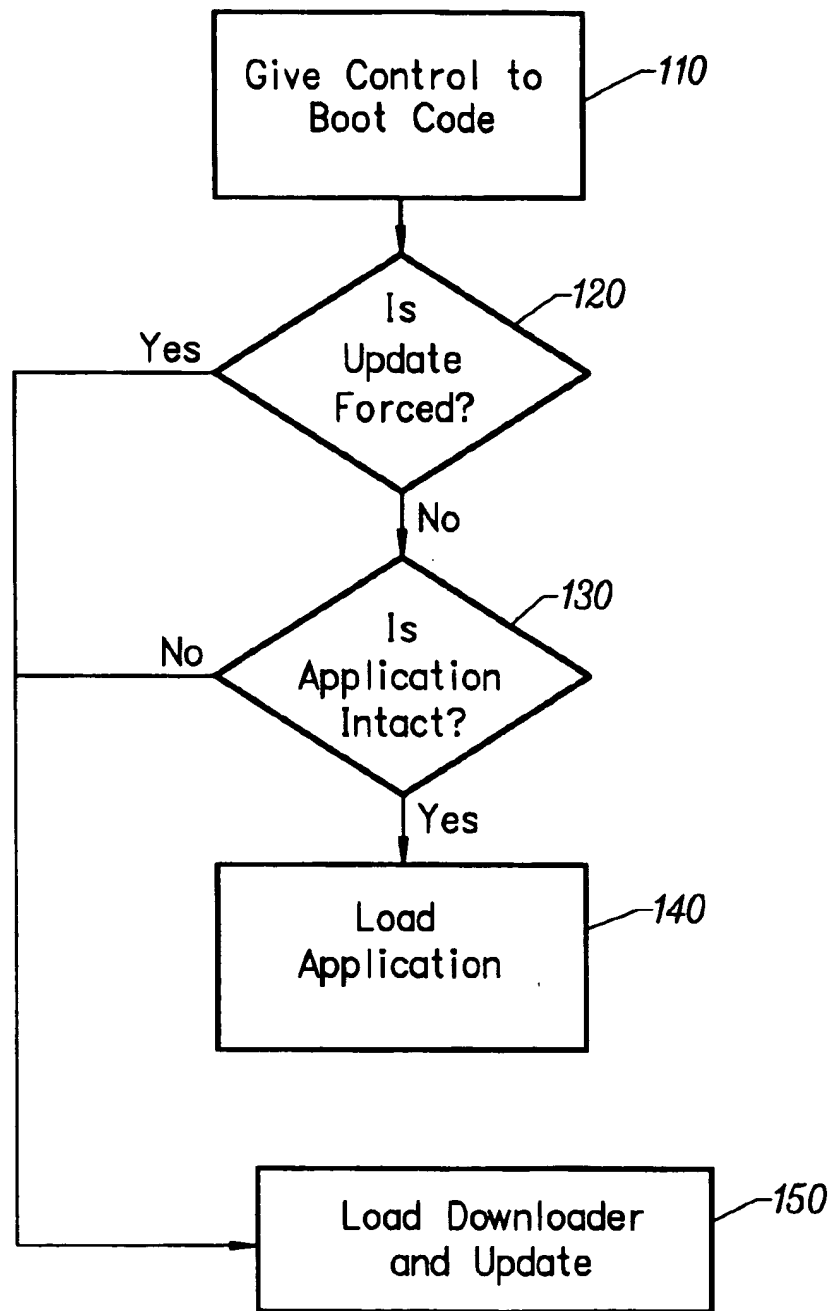


FIG. 1

*FIG. 2*

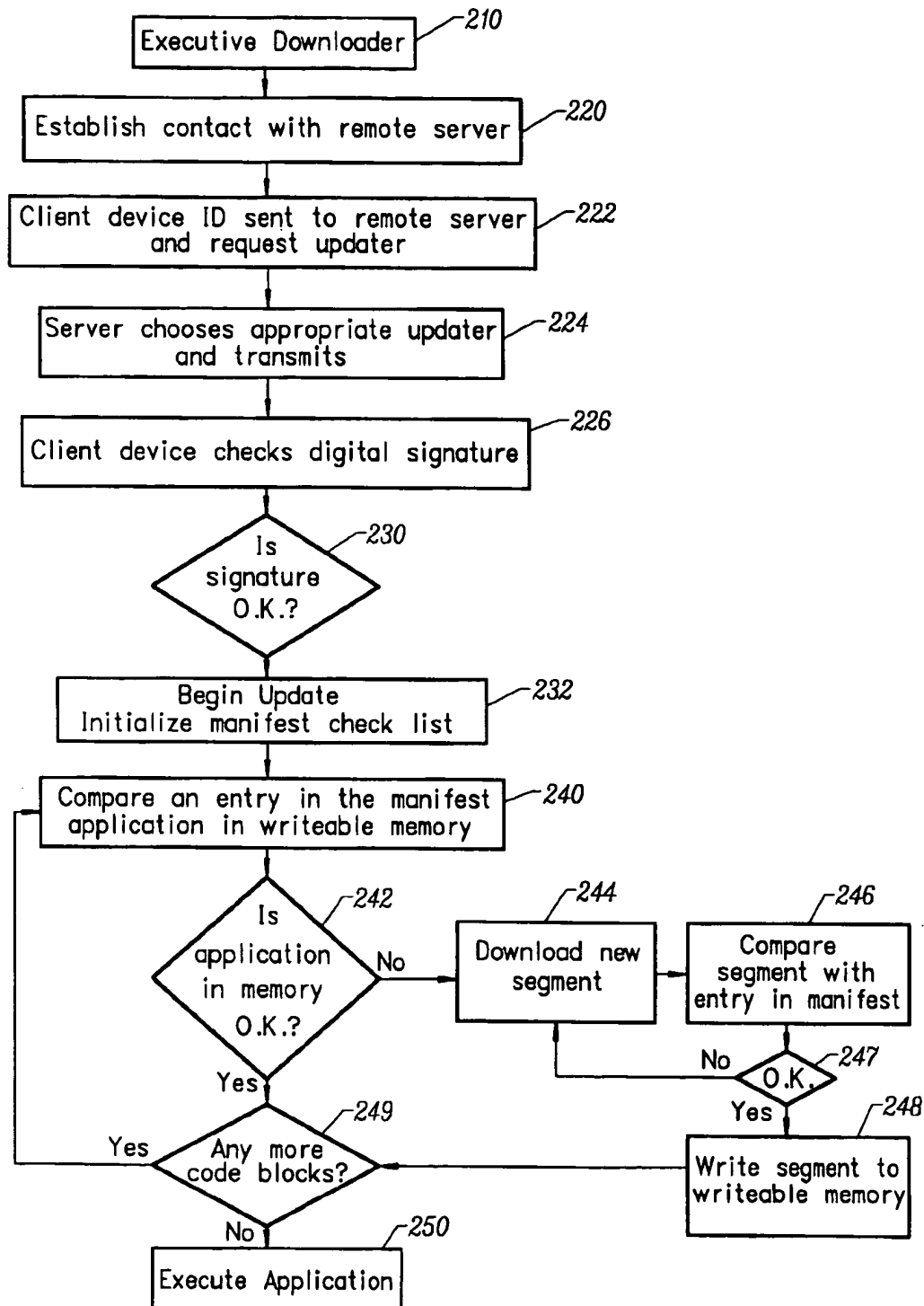


FIG. 3

1

SECURE DATA DOWNLOADING, RECOVERY AND UPGRADING

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation of application Ser. No. 09/080,577 filed May 18, 1998, now allowed; which is a continuation in part of application Ser. No. 08/770,238, filed Dec. 20, 1996, in the name of inventors Wei Yen and Steven Weinstein, titled "Internet Multiplexer for Broadcast and Other Information," now U.S. Pat. No. 5,991,799; which claims benefit of Provisional Application Serial No. 60/046,749, filed May 16, 1997, in the name of inventors Robert Shaw, Christopher Moeller, Clifford Mercer, Mark Law, and Luis Valente, titled "Operating System and Memory Management," now expired.

Each of these applications is hereby incorporated by reference as if fully set forth herein.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The invention relates to secure downloading, recovery and upgrading of data.

2. Description of Related Art

Recent developments in networking include client devices, which interact with a network to contact one or more server devices, and that are disposed for displaying information from those server devices. Division of responsibility among the clients and server allows each client device to use relatively fewer resources (such as processing power or memory) and therefore to be relatively inexpensive. Client devices can be manufactured en masse at relatively smaller cost and distributed to a large number of end users.

One problem in the known art is that client devices are subject to various failures. These can include hardware failures, which can damage software used to control the client device, and software failures, which can cause the client device to operate erroneously. It would therefore be advantageous to provide a method and system for recovery from memory errors in the client device. Moreover, there may be substantial upgrades to software designed for the client device developed after the client device has been manufactured and delivered to the end user. It would therefore be advantageous to provide a method and system for delivering these software upgrades to the client device.

This problem in the known art is exacerbated by several factors. First, as the client device is relatively inexpensive and within the complete physical control of the end user, it is unknown whether the software available at the client device can be trusted. Second, the client device itself cannot necessarily trust the data it receives from the server device it is coupled to if established over an insecure network, such as the internet. Third, the client device has relatively limited resources for communicating with the server device; in particular, the client device has relatively limited resources for rapidly receiving downloaded information from server devices.

Accordingly, it would be desirable to provide an improved method and system for secure downloading, recovery, and upgrading. This advantage is achieved in an embodiment of the invention in which a client device contacts a server device using a reliable software module. The reliable software module obtains trustworthy information with which to perform software and data integrity tests, and with which to locate data for recovery or upgrade of the client device.

2

SUMMARY OF THE INVENTION

The invention provides an improved method and system for secure downloading, recovery, and upgrading. A client device receives information from a server device using reliable software modules stored in permanent memory in the client device. The reliable software modules perform software and data integrity tests, and locate and retrieve data for recovery or upgrade of the client device. The client device confirms the trustworthiness of the received information device by comparing digital signatures or digests for the information it receives with known digital certificates in the reliable software module or received from known trusted server devices.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows a block diagram of a system for secure data downloading and upgrading.

FIG. 2 shows a process flow diagram for booting a client device and determining whether there is a need for downloading and upgrading.

FIG. 3 shows a process flow diagram of a method for secure data downloading and upgrading.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

In the following description, a preferred embodiment of the invention is described with regard to preferred process steps and data structures. However, those skilled in the art would recognize, after perusal of this application, that embodiments of the invention may be implemented using one or more general purpose processors (or special purpose processors adapted to the particular process steps and data structures) operating under program control, or other special purpose circuits, and that implementation of the preferred process steps and data structures described herein using such equipment would not require undue experimentation or further invention.

System Elements

FIG. 1 shows a block diagram of a preferred embodiment of a system for secure data downloading and upgrading.

A client device 10 has a processor 12 connected to a permanent memory 14 and writeable memory 16 by a data bus 18. Both permanent memory 14 and writeable memory 16 are non-volatile or other persistent memory. Permanent memory 14 is a non-writeable memory, such as a ROM, while the writeable memory 16 is a persistent memory, such as a NVRAM, flash memory or disk. Client device 10 also incorporates other memory, such as RAM into which code is loaded for execution.

Stored within permanent memory 14 is boot code 22, which controls the initialization boot process for client device 10 after a reset or power cycle. Boot code 22 is further subdivided into host boot code 30 and Navio boot code 32. The host boot code 30 is a vendor provided, client-resident code that if present initializes the system after power-on or a reset and then transfers control to the Navio boot code 32. The Navio boot code 32 is the code that is responsible the selection and execution of either the downloader code 24 or the application code 26.

The downloader code 24 is stored within permanent memory 14 and is a secure set of code which controls the connection of client device 10 via I/O port 40 to a communication channel 50. The downloader code 24 is the only copy of code that can be trusted after random or intentional corruption as it resides in permanent read-only storage. I/O

3

port 40 is connected to data bus 18, and permits data such as updater code 70 from a remote server 60 to be downloaded under the control of downloader 24 via communication channel 50 to writeable memory 16 (RAM). Updater code 70 is code that is stored on the remote server 60 and downloaded to the client device 10 to upgrade or recover the application code 26. The application code 26 is stored in writeable memory 16 and is the application seen by the user (i.e. the browser along with the operating system). In order to minimize the size of permanent memory 14 and to maximize flexibility in upgrading application code 26 for client device 10, it is desirable for boot code 22 to be as stream-lined as possible. Additionally, downloader 24 can be compressed in permanent memory 14 to save memory. Later it can be decompressed before being loaded into RAM for execution.

Communication channel 50 comprises a telephone line, ISDN line, cable, fiberoptic, or any other data transfer line connected to a modem or any other network interface device connected to, or contained within, I/O port 40. Such a connection links client device 10 through a network, either a private intranet or the public Internet to any of a plurality of remote servers 60.

Client device 10 includes a forced update feature 20 which when depressed reboots the client device 10. This forced update feature 20 does not need to be a physical switch. It may for example, consist of a combination of switches which when toggled in a specific sequence achieve a reboot and force a download. Alternatively, depressing a reset switch for a predetermined extended period of time, such as for five seconds, could activate the forced update feature 20.

Method of Operation

FIG. 2 shows a process flow diagram for booting client device 10 and determining whether there is a need for downloading and upgrading application code 26.

Initially, boot code 22, which is further subdivided into its host boot code and Navio boot code portions, takes control of client device 10. The host boot code 30, a vendor provided, client-resident code initializes the system after power-on or a reset and then transfers control to the Navio boot code 32. After control has been passed to the Navio boot code 32 the decision to run the application code 26 or the downloader 24 is made at step 110. The Navio boot code performs an integrity check on the application code 26 using a verification method such as an MD5 digest because at this part of the process only corruption is being checked as opposed to authenticity.

The writeable memory 16 stores two binary values as indicators of system integrity. The first is RunDownloader which if set causes the downloader 24 to run an upgrade of the application code 26. The second is TrustData which if not set creates the assumption that the data in writeable memory 16 is corrupt despite any appearance to the contrary. This is a validity check to determine whether the writeable memory 16 which contains information like the local ISP phone number and client-id/password as well as application code 26 may be trusted.

When either of these two indicators have these assigned values, a reset flag 28 is set in the writeable memory and the Navio boot code 32 forces a system reboot and update at step 120. Similarly, the user may also press the forced update feature and force a system reboot and update at step 120. In either case once the reboot and update has commenced, the downloader 24 initializes and connects to the remote server 60 and passes the information regarding client-id/password and synchronizes with the remote server 60 at which point the client device 10 may be given a new ISP phone number to call.

4

At step 130 it is determined whether the application code 26 is intact. If application code 26 is intact, control is transferred to application code 26 at 140 and client device 10 proceeds with its normal function. This is the usual result of the boot sequence. If application code 26 is corrupt, control of client device 10 is transferred to the update downloading sequence 150.

Retrieving the Updater

FIG. 2 illustrates the update downloading sequence of the updater 70 from remote server 60 at step 150. FIG. 3 illustrates the updating and downloading sequence in greater detail. Starting at step 210 where the decision has been made to load the downloader, the loading process continues at step 220 with the client device 10 establishing contact with the remote server 60. The client device 10 transmits at step 222 identification information regarding itself, e.g., version number, memory size, and requests transmission of an updater code package, a digitally-signed manifest which in some embodiments may contain additional program code for executing the update.

When the correct updater is located at step 224, a chunk of updater code 70 will be downloaded to the client device from the remote server 60. Next, client device 10 receives the transmitted updater code 70 and checks it at step 226 to determine whether it is valid. In the illustrated embodiment, downloader 24 compares a digital signature or digest contained within the updater package with known signature data saved in downloader code 24. Because downloader 24 is stored in permanent memory 14, it is secure and trustworthy, and hence a match between the digital signature of the updater code package and the stored digital signature validates the identity of the updater. This allows client device 10 to update application code 26 irrespective of the relative security of either communication channel 50 or remote server 60.

Next client device 10 reviews the results of the comparison at step 230. If the signature does not match, the process returns to step 220, preferably with connection to a different remote server 60, and a new updater is received. If the signatures match, the update process begins with an initialization process at step 232. The updater contains a table or manifest, which logically divides application code 26 into smaller code segments, representing portions of or complete logical segments of application code 26.

Retrieving the Application Code

When a valid updater is received the updater is given control of client device 10. It sets the "in update" flag, and loads a new update of code segment by segment. The updater checks each segment of application code 26 for corruption by determining whether the contents of a specific code segment match a hash or digest.

In a preferred embodiment this is implemented by including a manifest or table of secure hashes such as SHA1 in the updater 70. Each hash in the table should correspond to an image of the most recent segment of application code 26. Each hash in the manifest is accompanied by a location identifier, such as an internet URL, for a corresponding replacement code segment. This allows different replacement code segments to be stored in separate locations, such as different remote servers 60. This method allows individual components of application code 26 to be loaded separately.

After the initialization process at step 232, a segment of code from application 26 is verified at step 240 using the hash checklist. As shown at step 242, if that block of code is valid, client device 10 checks at step 249 to see if there are remaining code segments to be verified. If there are more

5

code segments, the process repeats starting at step 240 for the next code segment. If there aren't any more code segments to check, the application is loaded at step 250 and the update is complete.

If the block of code is invalid at step 242, a substitute segment of code is downloaded from the corresponding URL specified by the updater. The downloaded code segment may be encrypted to provide additional security or compressed for transmission to speed download time. After the code segment is downloaded at step 244 and any required decryption or decompression is performed, the new code segment is checked against its respective hash at step 246 and if it matches at step 247 it is then written to writeable memory 16 at step 248 to replace the corrupted segment of application code 26.

Launching the Application Code

When a hash does not correspond to an image of the most recent chunk of application code 26, the updater 70 downloads a file that is designated in the table. At step 246 a check is made to see if there are any more code segments of application code 26 to be checked. After all of the code segments of application code 26 have been checked and validated at steps 242 through 247 and any necessary replacement code segments written to memory at step 248, the application code 26 is completely loaded and ready to execute at step 250.

In a preferred embodiment each file downloaded by the updater 70 is compressed to minimize bandwidth and speed transmission through communication channel 50. In addition, each segment of code downloaded preferably has a default size of no greater than 64 kilobytes to correspond with the size of a sector of flash memory. Additionally, the segments of code to be downloaded should be in ascending order of memory address, non-overlapping and they should not cross a flash sector boundary.

Once a given segment of code is downloaded and decompressed, it is authenticated and validated by using the same hash that was used to check the segment of application code 26 which it is replacing and which originally failed the hash check. If the hash matches that in the table of updater 70 then the downloaded segment of code is authentic and is allowed to overwrite its respective segment of application code 26 in writeable memory 16. This is a means of linking security from the digital signature of the updater to the downloaded segments of code because the authentication happens in the process of downloading. Thus, the digital signature is on the downloaded datastream as well as the updater image in memory.

In a preferred embodiment the table of secure hashes, SHA1 or other equivalents thereof, has a dual purpose. Because the table of secure hashes can check the digest of segments of application code from the updater for validity without checking the whole code segment, a means of compression is achieved in addition to a validity check. For example, a twenty-byte long SHA1 hash can be used to verify a large code segment like a 128-kilobyte segment by skipping the segments that are valid. If only a few segments of code need to be replaced, the savings in transmission time and bandwidth will be substantial. This time savings is important not only to the user of client device 10, but also to the remote server 60 when a large number of client devices 10 need to be updated, such as when a new version of a segment of application code 26 is released or when the number of client devices in service is very large.

Using the process described above, the present invention ensures that a client device 10 can restore a corrupted application code 26 using minimal program code stored in

6

permanent memory 14. Because the validation process occurs each time client device 10 is initialized, the integrity of application code 26 can be maintained on an ongoing basis. Further, because of the use of digitally signed hashes in the updater, secure transmission of only those segments of application code 26 which need to be updated can be achieved using unsecured remote servers 60 and communication channels 50.

Alternative Embodiments

Although preferred embodiments are disclosed herein, many variations are possible which remain within the concept, scope, and spirit of the invention, and these variations would become clear to those skilled in the art after perusal of this application.

What is claimed is:

1. A method of updating application code on a client device from a remote server, said method comprising:

performing a boot sequence on said client device, said boot sequence including (a) verifying the validity of application software contained in a writeable memory, (b) automatically retrieving without user intervention from the remote server through a communication channel update data for identifying invalid code segments, (c) automatically retrieving without user intervention from the remote server through said communication channel replacement code for replacement of said invalid code segments, and comparing validity status data within said update data for identifying invalid code segments such that only invalid code segments need be replaced whereby a compression of data transmission is effected.

2. The method of claim 1, wherein said boot sequence further comprises checking for a forced update signal and initiating retrieval of said update data in response to said forced update signal.

3. The method of claim 1, wherein said boot sequence further comprises checking a forced update flag and initiating retrieval of said update data in response to said forced update flag.

4. The method of claim 1, wherein comparing said validity status data within said update data for identifying invalid code segments to be replaced is performed automatically without user intervention.

5. A system for updating application code from a remote server, said system comprising

a client device including a writeable memory;
an interface between said client device and a communication channel to said remote server;

software code stored in said writeable memory to perform a boot sequence on said client device, said boot sequence including (a) determining whether to update a software application stored in said writeable memory, and terminating the boot sequence if no update is necessary, (b) automatically retrieving without user intervention from said remote server over said communication channel update data for identifying invalid code segments, (c) automatically retrieving without user intervention from said remote server over said communication channel replacement code for replacement of said invalid code segments, (d) comparing validity status data within said update data for identifying invalid code segments such that only invalid code segments need be replaced whereby a compression of data transmission is effected, and (e) comparing authentication data within said update data with authentication data stored in said memory for authenticating said update data.

6. The system of claim 5, wherein step (a) further comprises checking for a forced update signal and initiating an update in response to said forced update signal.

7. The system of claim 5, wherein step (a) further comprises checking a forced update flag in said writeable memory and initiating an update in response to said forced update flag.

8. The system of claim 5, wherein said boot sequence further includes: (f) downloading and authenticating said replacement code from servers specified by location data within said update data upon successful authentication of said update data, and (g) writing said replacement code into said writeable memory.

9. The system of claim 8, wherein step (f) comprises determining from said validity status data in said update data which code segments of said software application require updating and downloading only those software code segments requiring updating.

10. The system of claim 8, wherein said validity status data comprises hash codes corresponding to discrete logical segments of code within said software application.

11. The system of claim 10, wherein said location data comprises internet addresses corresponding to said discrete logical segments of code.

12. The system of claim 8, wherein said replacement code is encrypted and step (f) further comprises decrypting said replacement code.

13. The system of claim 8, wherein said replacement code is compressed and step (f) further comprises decompressing said replacement code.

14. The system of claim 5, wherein said authentication data comprises a digital signature.

15. The system of claim 5, wherein comparing said validity status data within said update data for identifying invalid code segments to be replaced is performed automatically without user intervention.

16. A memory storing information including instructions, the instructions executable by a processor to update application code on a client device from a remote server, the instructions comprising

performing a boot sequence on said client device, said boot sequence including (a) verifying the validity of application software contained in said memory, (b) automatically retrieving without user intervention from the remote server through a communication channel update data for identifying invalid code segments, (c) automatically retrieving without user intervention from the remote server through said communication channel replacement code for replacement of said invalid code segments, and (d) comparing validity status data within said update data for identifying invalid code segments such that only invalid code segments need be replaced whereby a compression of data transmission is effected.

17. The memory of claim 16, wherein step (a) further comprises checking for a forced update signal and initiating retrieval of said update data in response to said forced update signal.

18. The memory of claim 16, wherein step (a) further comprises checking a forced update flag and initiating retrieval of said update data in response to said forced update flag.

19. The memory of claim 16, wherein comparing said validity status within said update data for identifying invalid

code segments to be replaced is performed automatically without user intervention.

20. A memory storing information including instructions, the instructions executable by a processor to update application code on a client device from a remote server over a communication channel, the instructions comprising:

performing a boot sequence under control of boot code stored in permanent memory in said client device, said boot sequence including (a) determining whether to update a software application stored in writeable non-volatile memory, and terminating the boot sequence if no update is necessary, (b) automatically retrieving without user intervention from said remote server over said communication channel update data for identifying invalid code segments, (c) automatically retrieving without user intervention from said remote server over said communication channel replacement code for replacement of said invalid code segments, (d) comparing validity status data within said update data for identifying invalid code segments such that only invalid code segments need be replaced whereby a compression of data transmission is effected, and (e) comparing authentication data within said update data with authentication data stored in said permanent memory for authenticating said update data.

21. The memory of claim 20, wherein step (a) further comprises checking for a forced update signal and initiating an update in response to said forced update signal.

22. The memory of claim 20, wherein step (a) further comprises checking a forced update flag in said writeable memory and initiating an update in response to said forced update flag.

23. The memory of claim 20, wherein said boot sequence further includes: (f) downloading and authenticating said replacement code from servers specified by location data within said update data upon successful authentication of said update data, and (g) writing said replacement code into said writeable memory.

24. The memory of claim 23, wherein step (f) comprises determining from said validity status data in said update data which code segments of said software application require updating and downloading only those software code segments requiring updating.

25. The memory of claim 23, wherein said validity status data comprises hash codes corresponding to discrete logical segments of code within said software application.

26. The memory of claim 25, wherein said location data comprises internet addresses corresponding to said discrete logical units of code.

27. The memory of claim 23, wherein said replacement code is encrypted and step (f) further comprises decrypting said replacement code.

28. The memory of claim 23, wherein said replacement code is compressed and step (f) further comprises decompressing said replacement code.

29. The memory of claim 20, wherein said authentication data comprises a digital signature.

30. The memory of claim 20, wherein comparing, said validity status data within said update data for identifying invalid code segments to be replaced is performed automatically without user intervention.

* * * * *